DOCUMENT RESUME

ED 452 825 IR 020 662

AUTHOR Heinecke, Walter F.; Blasi, Laura; Milman, Natalie; Washington, Lisa

TITLE New Directions in the Evaluation of the Effectiveness of Educational Technology.

PUB DATE 1999-00-00

NOTE 9p.; In: The Secretary's Conference on Educational Technology, 1999: Evaluating the Effectiveness of Technology. [Proceedings] (Washington, D.C., July 12-13, 1999); see IR 020 654.

AVAILABLE FROM For full text: http://www.ed.gov/technology/1999/whitepapers.

PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS Computer Assisted Instruction; Computer Uses in Education; *Educational Practices; *Educational Technology; Elementary Secondary Education; *Evaluation Criteria; Evaluation Methods; *Evaluation Needs; *Instructional Effectiveness

ABSTRACT

The Secretary's conference on evaluating the effectiveness of educational technology highlighted new and emerging data on technology effectiveness in primary and secondary education reflected in the latest research and promising practices. The intent was to influence the way educators, teachers, and policy makers evaluate and assess the growing investment in technology and to provide schools with tools and strategies for effective evaluation. This paper aims to inform the discussion by examining recent changes in evaluation theory and practices, and by clarifying some definitions of evaluation, technology and student learning. The paper highlights instances of promising practices and concludes with a list of recommendations concerning the evaluation of the effectiveness of technology in teaching and learning. These recommendations include the following. A more formative approach to the evaluation of technology is needed because of the rate of change in technologies. In order to get at the complexities of these processes, multiple measures (quantitative and qualitative) should be used. Evaluation design should incorporate longitudinal studies of cohorts of students over several years. In addition, evaluation designs should rely less on participants' self-reported attitudes and more on observations of participants' actions within learning contexts. Future evaluations should not focus on simple outcomes measures such as posttests, but should also focus on complex metrics describing the learning process, such as cognitive modeling. Implementation evaluations should be conducted prior to outcomes evaluations. Focus should be on description of the program, treatment or technological innovation, and stronger descriptions of how the technological innovation is configured should be developed. The complexity of educational technology should be recognized. (AEF)

Reproductions supplied by EDRS are the best that can be made from the original document.

ED 452 825

New Directions in the Evaluation of the Effectiveness of Educational Technology

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

By: Walter F. Heinecke, Laura Blasi, Natalie Milman, and Lisa Washington

IR020662

BEST COPY AVAILABLE

New Directions in the Evaluation of the Effectiveness of Educational Technology

Walter F. Heinecke, Ph.D, Laura Blasi
Natalie Milman, and Lisa Washington

Curry School of Education,
University of Viginia

Introduction

At the Secretary's conference on evaluating the effectiveness of educational technology we will be asked to address the following fundamental questions:

How does technology impact student learning?

What can we know about the relationship using data and tools available?

What can we learn about the relationship in the future with new tools and new strategies?

The conference will highlight new and emerging data on effectiveness of technology in primary and secondary education reflected in the latest research and promising practices. The intent of the proceedings is to influence the way educators, teachers, policy makers evaluate and assess the growing investment in technology and to provide schools with tools and strategies for effective evaluation.

In this paper we hope to inform the discussion by discussing recent changes in evaluation theory and practices, and by clarifying some definitions of evaluation, technology and student learning. It is evident that there are multiple definitions of evaluation, of technology and of student learning and theses multiple definitions must be engaged prior to substantive debate over the course of future directions. We will highlight what we believe are instances of promising practices and conclude with a list of recommendations concerning the evaluation of the effectiveness of technology in teaching and learning.

Recent Changes in Evaluation Practices

We should say at the outset that evaluation means many things to many people. According to Glass and Ellett (1980) "evaluation- more than any science- is what people say it is, and people currently are saying it is many different things" (cited in Shadish, Cook and Leviton, 1991, p. 30). In a recent examination of evaluation practice, we are encouraged to bring a critical eye to bear on the purpose and conduct of evaluations. Shadish Cook and Leviton (1991) recommend that in any evaluation endeavor we ask fundamental questions about five key issues:

1. **Social programming**: What are the important problems this program could address? Can the program be improved? Is it worth doing so? If not, what is worth doing?

   To maximize helpful change in the public interest, is it more effective to modify the philosophy or composition of whole programs, or to improve existing programs incrementally-perhaps by modifying regulations and practices, or influencing which local projects are phased out? Should the evaluator identify and work with change agents, or merely produce and explain evaluation results without forming alliances with change agents? Should evaluators try to change present programs or test ideas for future programs? Under what circumstances should the evaluator refuse to evaluate because the relevant problem is not very important or the problem is not likely to ameliorate the problem?

2. **Knowledge use**: How can I make sure my results get used quickly to help this program? Do I want to do so? If not, can my evaluation be useful in other ways?

   Should conceptual or instrumental use have priority? Should the evaluator identify and attend to intended users of evaluations? If so, which users? What increases the likelihood of use, especially for instrumental versus conceptual use?

3. **Valuing**: is this a good program? By which notion of "good"? What justifies the conclusion?

   By whose criteria of merit should we judge a social program? Should prescriptive ethical theories play a significant role in selecting criteria of merit? Should programs be compared to each other or to absolute standards of performance? Should results be synthesized into a single value judgment?

4. **Knowledge construction**: How do I know all this? What counts as a confident answer? What causes that confidence?

   How complex and knowable is the world, especially the social world? What are the consequences of oversimplifying complexity? Does any epistemological or ontological paradigm deserve widespread support? What priority should be given to different kinds of knowledge, and why? What methods should evaluators use, and what are the key parameters that influence that choice?

5. **Evaluation practice**: Given limited skills, time, and resources, and given the seemingly unlimited possibilities, how can I narrow my options to do a feasible evaluation? What is my role-educator, methodological expert, judge of the program- worth? What questions should I ask, and what methods should I use?

   What should the role of the evaluator be? Whose values should be represented in the evaluation? Which questions should the evaluator ask? Given limited time and resources, which methods should be used to best answer the questions? What should the evaluator do to facilitate use? What are the important contingencies in evaluation practice that guide these choices?

Experts on program evaluation (House, 1993; Schorr, 1997; Shadish, Cook and Leviton, 1991) all indicate that program evaluation has undergone a major transformation in the last three decades. It has changed from "monolithic to pluralist conceptions, to multiple methods, multiple measures, multiple criteria, multiple perspectives, multiple audiences, and even multiple interests. Methodologically, evaluation moved from primary emphasis on quantitative methods, in which the standardized achievement test employed in a randomized experimental control group design was mostly highly regarded, to a more permissive atmosphere in which qualitative research methods were acceptable (House 1993. p. 3) The most fundamental shift has been away from a blind faith in the science of

evaluation and experimental research methods based on standardized test scores. These changes in the practice of evaluation have significant implications for questions about the future of the evaluation of technology and student learning outcomes.

The primary question to which we always turn to is: How does technology impact student learning? We don't, however, make implementation decisions based on this question. What do we know about this relationship using data and evaluation tools currently available and what could we learn in the future about technology and student learning assuming the application of new evaluation tools and strategies? The answer to the first question is fairly straight forward: The relationship depends on how you define student learning and how you define technology.

If one defines student learning as the retention of basic skills and content information as reflected on norm referenced and criterion referenced standardized tests, then, evidence suggests, there is a positive relationship between certain types of technology and test results. For instance, it is well established that if a teacher uses computer assisted instruction or computer based learning approaches, where the computer is used to manage the "drill and skill" approach to teaching and learning, students will show gains on standardized test scores. This view of technology reduces the equation to only a student, a computer and a test. It ignores the effects of schools, teachers, and family and community life on the learning process. Even tough we cannot control for these variables, we must not discount them.

If, on the other hand, one views the goal of education as the production of students who can engage in critical, higher order, problem-based inquiry, new potential for entirely different uses of technology emerge. For instance, the world wide web can be used as a source of information from which students can draw to solve real world problems by applying technology knowledge and skills. We can evaluate these outcomes but it is more complicated than the standardized testing route. Standardized tests are an efficient means for measuring certain types of learning outcomes but we must again ask ourselves, are these the outcomes we value for the new millennium? To a certain extent we are living out the decisions reflected in previous evaluation methods which constrain our thinking about the purpose and effectiveness of technology in education.

Policymakers, evaluators and practitioners may have vary different answers to fundamental questions about the effectiveness of educational technology. Everyone is asking for results of the investment of technology in education. Perhaps the primary difficulty in coming up with new ways of evaluating or assessing the impact of education technology is that there is little consensus about its purpose (Trotter, 1998). Policy makers often work from a cost-benefit model with increases in norm referenced and criterion referenced test scores viewed as the primary benefits. This appears to be at odds with the view held by teachers or by the public that educational technology benefits include preparing students for jobs, increasing student interest in learning, increasing student access to information and making learning an active experience (all rated above technology's impact on basic skills by parents in a 1998 public opinion survey sponsored by the Milken Exchange).

The question really should not be does educational technology work? "but when does it work and under what conditions?" (Hasselbring cited in Viadera, 1997). In practice, student achievement outcomes are mediated by the processes of teacher integration of technology into instruction. Technology can be used to improve basic skills through automated practice of drill and skill. Technology can also be used to facilitate changes in teacher practices that promote critical, analytic, higher order thinking skills and real-world problem solving abilities by students. The ability of teachers to foster such changes depends

significantly on training that shows them how to integrate technology into content specific instructional methods. This has been shown through programs such as the Adventures of Jasper Woodbury conducted at Vanderbuilt University, the national Geographic Society's Kid's network, and work done at University of Massachusetts, MIT and TERC with Simcalc.

Any innovation in our system of education, including technology, raises persistent questions about the purposes of education. Is it to provide training in fundamental and basic skills? Is it to prepare students for the work force? Is it to produce citizens for an effective democracy? Is it to produce an equitable society? Is it to produce broad, life-long learners? Is it to prepare students with critical thinking skills for a complex new world? According to educational researcher Larry Cuban, unless educational policy makers can agree and clarify the goals for using technology, it makes little sense to try and evaluate it.

This raises questions about assessment and evaluation of educational technology. Do traditional, standardized assessments measure the benefits that students receive from educational technology? In the evaluation of social programs in general, the profession of evaluation has moved away from standardized test scores as a meaningful measure of the impact of programs. Evaluation theorists like Mackie and Cronbach have argued that there are too many critical relationships occurring in social phenomenon to be adequately captured by the traditional experimental design. "Social programs are far more complex composites, themselves produced by many factors that interact with one another to produce quite variable outcomes. Determining contingent relations between the program and its outcomes is not as simple as the regulatory theory posits" (House, 1993, p. 135-6). Besides improvements in retention of rote facts, technology can improve student attitudes toward the learning process. perhaps we should be assessing actual, authentic tasks produced through the processes of student interaction and collaboration. Perhaps we should be developing technologically based performance assessments to measure the impact of technology on student learning.

We have been fairly successful in determining the impact of technology on basic information retention and procedural knowledge. However, we have been less than successful in evaluating the impact of educational technology on higher order or metacognitive tinning skills.

Needed: New and Expanded Definitions of Student Learning Outcomes

What are needed more than anything else are a new set of clear learning outcomes for students who must live in a complex world. New learning outcomes must focus on the demands of the new world environment. We need students who can think critically, solve real world problems using technology, take charge of their life-long learning process, work collaboratively and participate as citizens in a democracy. Experts in the area of technology and education such as Jan Hawkins and Henry Becker have provide ideas that could be developed into criteria for new ways of thinking about technology, teaching and learning. These new learning outcomes could be translated into learning benchmarks and new types of assessment and methods for measuring outcomes could developed to measure these benchmarks.

What we are looking for is a transition from isolated skills practice to integrating technologies as tools throughout the disciplines. Jan Hawkins argued that to realize high standards, education needs to move beyond traditional strategies of whole group instruction and passive absorption of facts by students. New more effective methods are based on engaging student in complex and meaningful problem-solving tasks. Technologies need to be used to bring vast information resources into the classrooms. We need a transition from inadequate support and training of teachers to support for all teachers to learn how to use

technologies effectively in everyday teaching (Hawkins, 1996).

According to Becker (1992) in an ideal setting, teachers use a variety of computer software, often working collaboratively to address curricular goals. Students exploit intellectual tools for writing, analyzing data, and solving problems and they become more comfortable and confident about using computers (Becker, p. 6). Exemplary teachers use computers in lab settings as well as classroom settings at the school for consequential activities that is where computers are used to accomplish authentic tasks rather tan busywork such as worksheets, homework assignments, quizzes or tests. Means and Olson (1994) outline a set of criteria for successful technology integration projects: An authentic challenging task, a project where all students practice advanced skills, where work takes place in a heterogeneous, collaborative groups, the teacher acts as coach and provides guidance, and where work occurs over extended blocks of time..

Evaluating for New Visions of Technology Teaching and Learning

It is clear that teaching and learning processes are embedded within complex systems. The challenge is to develop evaluation models that reflect this complexity. Just as technology has caused us to reevaluate the nature of knowledge and instruction, it prompts us to reevaluate the forms of evaluation that are brought to bear when examining educational technology. According to Schorr (1997) we need a new approach to the evaluation of complex social programs, one that is theory-based, aiming to investigate the project participant's theory of the program; one that emphasizes shared rather than adversarial interests between evaluators and program participants; one that employs multiple methods designs; and, one that aims to produce knowledge that is both rigorous and relevant to decision-makers. In order to accomplish these tasks it will be necessary to design evaluations of technology in K-12 settings based on the experiences of evaluators, the experiences of program developers, "state of the art" in the field of technology and learning and the various program descriptions.

Several studies and reports have done an exemplary job at pointing us in promising directions for future evaluations of the effectiveness of educational technology. For instance Bodily and Mitchell have prepared an evaluation sourcebook for "Evaluating Challenge Grants for Technology in Education" published by the RAND Corporation. Bodilly and Mitchell (1997) acknowledge that the outcomes sought in technology infusion projects are complex and "not entirely captured by traditional educational measures, seeking better learning outcomes "on a complex variety of dimensions rather than improvements in traditional test scores" but they go on to recommend that some stake holders may be interested in test scores as measures of student learning. They indicate that performance outcomes are the results of complex causes. Technology may be only one of many input variables causing changes. A project's implementation and outcomes are heavily influenced by its context. Goals of various educational technology projects are unique and may not be captured by a uniform evaluation design and multiple evaluation design are required.

In terms of outcome goal, they include a wide variety of possibilities beyond traditional test scores including: short term changes in student outcomes like disciplinary referrals, homework assignments completed or longer term indicators such as changes in test scores or student performances, increased college going rates, increases in job offers to students. Other outcomes are defined as higher order thinking skills, more sophisticated communication skills, research skills, and social skills. More sophisticated outcome measures must be located or developed by evaluators in order to gauge new effects of technology on learning.

Other outcome measures might be found in participants' (teachers and students) perceptions about the implementation, quality and benefits of the program. These might reflect student engagement levels as well as satisfaction levels. Other interim performance indicators might include the effect of the pro0gram on community and family participation or involvement, and student and teacher retention. Declines in disciplinary referrals and special education placements may also serve as outcome measures. The federal government, state departments of education, school district or schools might develop criteria for standards of good practice indicators and associate learning outcome benchmarks.

Other indicators of student outcomes such as higher order thinking skills and ability to apply knowledge in meaningful ways might be measured by performance assessments, portfolios, learning records, and exhibitions. Of course norm referenced and criterion referenced assessments can also supplement these alternatives outcomes. School districts are encouraged to use multiple and varied measures of outcomes. Student performance indicators such as attendance, reductions in drop-out rates, successful transitions to work and post-secondary institutions should be considered. Baseline data should be established at the beginning of the project. They also propose that a list of common indicators across projects be used as a tool for summative program evaluation.

Bodilly and Mitchell refer to work on the evaluation of technology in educational reform conducted by Herman (1995) and Means (1995). They conclude that broad-based technological reforms, those that attempt multiple changes in a school besides the insertion of a single computer-based course, such as an attempt to create a constructivist curriculum across all grade levels supported by computer technology are more difficult to measure in terms of outcomes. They state: efforts to trace the effects of these projects must take into account measuring effects in dynamic situations where many variables cannot be controlled and where interventions and outcomes have not been well defined for measurement" (p. 16). They also assert: "The complex environments in which technology projects are embedded make inference of causal relations between project activities and outcomes tenuous" (p. 20).

Implementation analysis becomes important under these conditions. With all of these complexities, effects of technology on student outcomes may not occur in the short-term evaluations must take into account the different phases of a schools integration of technology: purchasing and installing hardware and software, training teachers, integrating technology into the curriculum and instruction. Evaluation designs must therefore, be longitudinal in design and account for changes in the target population. Tracking comparison groups not exposed to technology or using national surveys to assess the likely level of background effects will often be necessary.

CMC corporation conducted a two year evaluation of the Boulder Valley Internet Project. The project employed a variety of evaluation method and developed a theoretical tool, The Integrated Technology Adoption Diffusion Model, to guide the evaluation. Evaluations should include the contexts within which technological innovations occur. This includes looking at technological factors, individual factors, organizational factors and teaching and learning issues (See Sherry, Lawyer-Brook, and Black, 1997). Evaluation designs must be flexible enough to attend to the varying degrees of adaptation occurring with different content areas. Evaluations must include implementation assessments, formative assessments as well as standard summative and outcomes assessments. Evaluations must include the quality of training programs offering teachers the opportunity to learn new technologies within relevant, subject-specific contexts.

Recommendations

We need to take a more formative approach to the evaluation of technology because of the rate of change in technologies. Technology changes so quickly that teachers are often asked to keep up and integrate new ideas at break neck speeds. The definition of what is the innovation is thus constantly at issue and we must spend time documenting the program which may be changing over time..

In order to get at the complexities of these processes multiple measures (quantitative and qualitative) should be used. These should include traditional experimental and quasi-experimental designs and include such methods as paper surveys, email/web-based surveys, informal and in-depth interviews, focus group interviews, classroom observations and document analysis.

Evaluation design should incorporate longitudinal studies of cohorts of students over several years. In addition evaluation designs should rely less of participants self reported attitudes and more on observations of participants actions within learning contexts. We need to be in classrooms to observe how teachers are incorporating technology into their instruction and what effect this is having on student learning processes. We would recommend further efforts such as those by Milken and Elliot Soloway, to improve the format for research designs to allow for comparisons across sites.

Future evaluations should not focus on simple outcomes measures such as posttests but should also focus on complex metrics describing the learning process such as cognitive modeling (Merrill, 1995). Research and evaluation needs to demonstrate the potential of educational technology but in a way that attends to the layers of complexity that surround the processes. We need to include a wide variety of experts and stakeholders.

Conduct implementation evaluations prior to outcomes evaluations. Spend time necessary to determine whether an innovation as been adopted or fully implemented before trying to determine its effectiveness.

Focus on description of the program, treatment, or technological innovation, develop stronger descriptions of how the technological innovation is configured.

Recognize the complexity of educational technology; Define technology as an innovative process linking teaching and learning outcomes rather than a product which is dropped into the black box of teaching and learning outcomes defined as improvements on standardized test scores. Reduce the reliance on standardized test scores as the primary evaluation outcome. Replace dogmatic applications of experimental designs with designs that allow us to view the complexity of technology based reforms of teaching and learning from multiple perspectives. Adopt multifaceted approaches to evaluation that include case studies and theoretical modeling which includes individual, organizational, technological and teaching/learning aspects of adoption and diffusion of innovations. This means that participant observation of programs will be used as a form of data collection. This type of data collection is not inexpensive but provides evidence beyond self reported data or gross outcome measures like test scores.
-###-

[The Cyberspace Regionalization Project:]

[Measurement Issues with Instructional and Home Learning Technologies]